Unsupervised Cross-domain Image Generation with Generative Adversarial Network

Liwei Cai 2014011255
Department Electronic Engineering
Tsinghua University
clw14@mails.tsinghua.edu.cn

Kaidi Cao 2014012282
Department Electronic Engineering
Tsinghua University
ckd14@mails.tsinghua.edu.cn

Abstract

To be fulfilled.

1 Introduction

Recently, generative adversarial networks (GAN) have drawn great attention from deep learning researchers. As this course focuses on "advanced application", we are also interested in learning about GAN. In our project, we decided to follow a fresh new publication [1] on generating images from unlabeled cross-domain data using GAN, which we considered novel, interesting, and not too complicated. We plan to reproduce and improve its result as far as we can.

[1] performs two experiments, as two examples of cross-domain image generating: generating handwritten digits (as in MNIST) from image of street view containing digits, and generating emoji style avatar from image of human face. Speaking in a general and mathematical manner, cross-domain sample generating tries to learn a mapping $G : S \rightarrow T$ between two domains S (source) and T (target), given a predetermined measuring function f, to minimize the difference of $f(x)$ and $f(G(x))$ for $x \in S$.

2 Related Works

2.1 GAN

The general idea of GAN, as proposed in [2], is to train a generator network along with a discriminator network, which tries to tell whether the sample is "real" (from dataset) or "fake" (generated by the network). The target is to make the generator "fools" the discriminator best, that is, to minimize the accuracy of the discriminator. Using adequate metrics and regularizing constraints, a generator network that can fool the discriminator can also fool human.

Original GAN can only generated random sample from the whole space, while a modified version called conditional GAN [3] can generated samples of a specific class or satisfying certain constraints. In this case, the condition would be taken as an input of the discriminator, and the output of the discriminator would be ternary, telling not only whether the generated sample is "real" but also whether it meets the condition.

[6] try to understand the machanism of GAN using information theory. It introduces a modified version of GAN, by maximizing the mutual information between a fixed small subset of the GAN's noise variables and the observation.

[7] proposes another version by combining variational autoencoder(VAE) with a GAN. In this way, one can use learned feature representations in the GAN discriminator as basis for the VAE reconstruction objective. This makes it possible to replace element-wise errors with feature-wise errors, in order to better capture the data distribution.

Inspired by the prominent performance on semi-supervised application of GAN, [8] introduced another image synthesizing model named Auxilary Classification GAN(AC-GAN), aiming at producing more discriminable images, AC-GAN was able to generate globally coherent samples(like improve the visual appearence of 128×128 images).

2.2 Neural style

Neural style transfer can be treated as a special case of cross-domain image generation. It transfer the style of one image (usually a painting) onto another image , without the need of any extra information (thus also being an unsupervised model), by performing gradient descent on the generated image to minimize the difference of certain features extracted by a CNN between itself and both inputs. [4] is a representative work of it, and have gained great public attention.

The core idea of [4] is to generate a new image using optimization method, which makes the generating process unavoidably slow. [5] combined the idea of optimization and feed-forward network, designed a network which is capable of generating styled images using simple forward method.

3 Plan of further works

Our plan is to closely follow the work of [1]. First, we will reproduce the result of [1], using exactly the same network structure. Then, we will try to improve it, by modifying network parameters, or by changing or even redesigning the network structure, depending on the amount of available time.

We have just determined on this topic a few days ago, and currently no experiment had been carried out.

References

[1] Yaniv Taigman, Adam Polyak & Lior Wolf (2016). Unsupervised Cross-domain Image Generation. ICLR 2017.

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in Neural Information Processing Systems (pp. 2672-2680).

[3] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. ICML 2016.

[4] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2414-2423).

[5] Johnson, J., Alahi, A., & Li, F. F. (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution. Computer Vision ECCV 2016. Springer International Publishing.

[6] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). Infogan: interpretable representation learning by information maximizing generative adversarial nets.

[7] Larsen, A. B. L., Sønderby, S. K., Larochelle, H., & Winther, O. (2016). Autoencoding beyond pixels using a learned similarity metric.

[8] Odena, A., Olah, C., & Shlens, J. (2016). Conditional image synthesis with auxiliary classifier gans.